



Round, J., & Hawton, A. (2017). Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values. *PharmacoEconomics - Open*. <https://doi.org/10.1007/s41669-017-0027-2>

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC

Link to published version (if available):
[10.1007/s41669-017-0027-2](https://doi.org/10.1007/s41669-017-0027-2)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer at DOI: 10.1007/s41669-017-0027-2. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values

Jeff Round¹  · Annie Hawton^{2,3}

© The Author(s) 2017. This article is an open access publication

Abstract Mapping between non-preference- and preference-based health-related quality-of-life instruments has become a common technique for estimating health state utility values for use in economic evaluations. Despite the increased use of mapped health state utility estimates in health technology assessment and economic evaluation, the methods for deriving them have not been fully justified. Recent guidelines aim to standardise reporting of the methods used to map between instruments but do not address fundamental concerns in the underlying conceptual model. Current mapping methods ignore the important conceptual issues that arise when extrapolating results from potentially unrelated measures. At the crux of the mapping problem is a question of validity; because one instrument can be used to predict the scores on another, does this mean that the same preference for health is being measured in actual and estimated health state utility values? We refer to this as conceptual validity. This paper aims to (1) explain the idea of conceptual validity in mapping and its implications; (2) consider the consequences of poor conceptual validity when mapping for decision making in the context of healthcare resource allocation; and (3) offer some

preliminary suggestions for improving conceptual validity in mapping.

Key Points for Decision Makers

Estimating health state utility values using mapping algorithms is becoming commonplace.

The conceptual basis of this process is poorly developed; that one instrument can be used to predict the scores on another does not mean that the same preference for health is being measured.

Decision validity is contingent on health state utility values meaning what is claimed.

1 Introduction

Mapping (sometimes called ‘cross-walking’) has become a common technique for estimating health state utility values for use in economic evaluations [1]. It typically refers to the process of statistically estimating health state utility values for use in an economic evaluation by predicting results for a target preference-based measure (e.g. the EQ-5D [2]) from data collected using a non-preference-based measure (often a condition-specific measure) [3]. The publication and use of mapping algorithms has increased significantly since, in its methodology guidelines for technology appraisals, the National Institute for Health and Care Excellence recommended the use of mapped health state utility estimates when data collected directly from patients are not available [4, 5].

✉ Jeff Round
jeff.round@bristol.ac.uk

¹ School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK

² Health Economics Group, University of Exeter Medical School, University of Exeter, Exeter, UK

³ Peninsula Collaboration for Leadership in Applied Health Research and Care, University of Exeter Medical School, University of Exeter, Exeter, UK

A large number of non-preference-based instruments have been, and continue to be, developed for use in health research. In particular, condition-specific, non-preference-based measures represent a rich source of information about the quality of life of studied populations. However, because no set of preference weights exist for these measures, they have limited use in allocating scarce healthcare resources, not only between treatments for the same or similar conditions but also especially across different areas of healthcare. The development of a statistical model that enables researchers to translate information gained from a non-preference-based instrument into health state utility scores for use in economic evaluations has use in a variety of scenarios. For instance, such mapped estimates allow data to be used to evaluate technologies where a condition-specific measure of quality of life has been collected during an evaluation, but not a preference-based measure. Yet despite the increased level of interest in mapping [1, 3, 6] and the publication of reporting guidelines [7], the validity of using mapping to estimate utility values has not been fully addressed.

McCabe et al. [8] identified multiple areas of concern in the development of mapping algorithms. In particular, they argued that it cannot be assumed that utility values predicted by a mapping algorithm are representative of directly measured utility values. Therefore, at the crux of the mapping problem is a question of validity. Just because one instrument can be used to predict the scores on another, does this mean that the same preference for health is being measured in actual and estimated health state utility values? In this article, we describe this as the problem of ‘conceptual validity’ [9]. We distinguish conceptual validity from the related concept of content validity. The latter is related to how well a single instrument reflects the underlying constructs being measured. On the other hand, conceptual validity relates to the degree to which the content of two different instruments reflect one another when used for mapping.

This paper aims to (1) explain the idea of conceptual validity in mapping and its implications; (2) consider the consequences of poor conceptual validity when mapping for decision making in the context of healthcare resource allocation; and (3) offer some preliminary suggestions for improving conceptual validity in mapping.

2 Mapping and Conceptual Validity

Current methods of mapping have focused on the mean or average accuracy with which the results of one instrument can be predicted based on responses to a second instrument [3, 10]. Mapping techniques have become increasingly statistically sophisticated [11, 12] and accepted approaches

for demonstrating validity are broadly statistical [4]. The final function is determined based on performance according to some error measurement statistic, typically either mean absolute error (MAE) or root mean squared error (RMSE). By finding a model that, on average, best estimates actual scores, it is generally assumed that this is therefore the model that leads to the greatest reduction in uncertainty in the final mapped estimates.

2.1 Transmutation of Data in the Mapping Process

Nonetheless, against this statistical backdrop, the conceptual context of what happens when mapping occurs seems to have been lost, as has consideration of what the mapped estimates mean. It is not enough to be able to say that the scores of Measure A are highly predictive of scores for Measure B if it cannot be said with any confidence that A and B are measuring the same thing. Estimating a reliable mapping function using an appropriate statistical method is a necessary condition for a mapping process to be considered valid, but it is of itself not sufficient.

Preference-based generic instruments such as the EQ-5D [2] and the SF-6D [13, 14] have been developed to give a broad picture of overall health but, as a result, have been considered insensitive to small changes in specific aspects of health. Non-preference-based measures of quality of life can be mapped to preference-based measures to provide estimates of utility scores. These may cover broadly similar domains. In some cases, they may be too lengthy for the development of a preference set (e.g. the WHOQOL [15]). In the event that the non-preference-based measure is itself brief, it may be subject to the same criticism as brief preference-based measures in that it may not be sensitive to change in the study population.

To address concerns around the sensitivity to change of generic instruments, condition-specific instruments are often used. Condition-specific instruments, such as the Asthma Quality of Life Questionnaire (AQLQ) [16], measure different domains of health than generic instruments, or measure the same domains but at a different level of detail. These measures are often designed to be more sensitive to changes in health that are specifically due to that condition of interest, although they may not be good indicators of overall health or preferences for health [17]. It is, in part, these differing objectives in instrument design that may lead to a lack of conceptual validity in mapped estimates.

Mapping methods also often include consideration of the clinical and demographic overlap between the estimation sample and the target sample to provide more robust results. However, including covariates in mapping functions can reduce their functionality. Where covariates are an important part of the prediction function, the end user in

turn needs to have details of the same covariate variables for their study participants in order to use the algorithm. As such, covariates are not always included in mapping functions. In addition, the nature of the covariates in the estimation and target samples does not affect the extent to which the two instruments are measuring the same construct.

2.2 Changing Tin into Gold?

Conceptual validity is important to consider when assessing the degree to which the source and target instruments are measuring the same domains of health. No two instruments will cover the exact same domains in the exact same fashion, but it is important to consider the degree of conceptual coherence between the source and target data. Where there is little overlap between the constructs of health being measured, mapping functions are unlikely to be valid [18]. Consider the EQ-5D-3L [2], the most commonly used preference-based target instrument in mapping studies [1, 3]. The EQ-5D-3L assesses five domains of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Other instruments, typically measuring different health domains, are often mapped to the EQ-5D preference set. For example the AQLQ [16] measures just four domains of health considered to be specifically relevant to patients with asthma: symptoms, activity limitation, emotional function and environmental stimuli. While there is some overlap with the EQ-5D, it is clear that the two instruments do not measure exactly the same aspects of health. As utility values derived from the EQ-5D have an upper and lower limit, accounting for the information content in the AQLQ implicitly means that greater preference will be given to the respiratory domain at the expense of other domains; however, it is not currently possible to quantify the extent to which this occurs.

Many studies use such condition-specific measures as the source instrument when mapping to a preference-based generic measure [1]. Condition-specific measures are often narrower or more focussed in scope (by design) than generic preference-based instruments, meaning that more information may be collected, but pertaining to a narrower set of domains of health. This additional information will be lost when mapping between instruments, and, more importantly, potentially relevant information will be missed relating to those domains of the preference-based instrument that are not captured in the condition-specific measure. This also has the negative effect of not clearly accounting for potential interaction effects in the domains being measured in mapped estimates, which amounts to an unaccounted loss of information.

It can be argued that where an algorithm has a sufficient prediction accuracy, use of a mapping algorithm to

estimate utility values should be based on whether the algorithm will consistently reflect the relationship between the instruments in an external population. This would reflect a high degree of external validity and could provide sufficient justification for use of the algorithm. However, measures of error in the prediction algorithm (MAE, RMSE) do not tell us what information comprises these estimates and therefore, for example, how they are likely to perform in relation to treatment effects. If one set of EQ-5D mapped values are estimated from scores on a walking scale and another set of EQ-5D mapped values are estimated from scores on a fatigue scale, using the mapped values in a cost-effectiveness analysis is likely to result in different estimates of cost effectiveness if the intervention focuses on improving mobility or improving fatigue. This could be tested empirically.

In support of mapping, it can also be argued that the mapped estimates reduce variation around the mean estimate compared with directly observed values [18]; however, this is misleading. First, it does not address the content problem, i.e. that the source and target instruments are measuring similar constructs of health, and, second, it says nothing about what information is lost during the mapping.

2.3 Transmuting Lead to Gold?

This loss of information is only likely to be exacerbated when mapping from a clinical measure to estimate health state utility values (e.g. [19–21]) as is allowable under National Institute for Health and Care Excellence (NICE) guidance, given that an adequate mapping function can be demonstrated and validated [4]. Clinical indicators capture particular aspects of health status and do not necessarily capture important changes to health-related quality-of-life (HRQoL) that may occur as a result of treatment. For example, an improvement in mobility that enables someone to be able to walk upstairs may have little impact on their HRQoL if they live in a bungalow, or a small improvement in pain may have a significant impact on an individual's self-reported HRQoL [17]. Using clinical indicators as the source data to map to a preference-based measure therefore limits the amount of information that can be translated into health state utility scores. In effect, this fails to overcome the principle objection to using clinical indicators, which is that comparisons cannot be made between interventions for different conditions because preferences for health states cannot be known from clinical indicators alone. A simple mapped estimate of preferences does not address this problem. Such mapped estimates will reflect only those domains directly relevant to the indicator selected, without capturing broader information about HRQoL. Because mapped estimates based solely on

clinical outcome indicators cannot distinguish which aspects of HRQoL are being captured in the final estimate, it is unlikely that clinical indicators can ever be a valid source for mapped estimates.

2.4 Turning Silver into Gold?

A third key problem relating to conceptual validity exists where the same domains of health are being measured. This arises from different ways of measuring the severity of each problem in a given domain of health, including those cases where there is a clear overlap between instruments in terms of the domains they assess. For example, the SF-12 [13, 14] and the EQ-5D both include dimensions regarding mental health (in the SF-12 this is the mental health component score, and in the EQ-5D this is the anxiety/depression dimension), but they give different amounts of information. The SF-12 includes six questions that relate to mental health, compared with just one question in the EQ-5D. Mapping from the SF-12 to the EQ-5D will inevitably lead to a loss of information pertaining to this domain.

When a non-preference-based measure of health outcome is mapped to a preference-based measure, it is not known what domains of health or HRQoL are retained in the resulting estimates and those aspects that are lost. The nature of the preference for health that is lost in mapped estimates is unknown, as is the nature of the preference for health that is retained. This raises questions about the validity of mapped estimates for use in healthcare decision making, as described in Sect. 4.2.

The problem of conceptual validity, or even the loss of information during the mapping process has not been fully addressed in the literature. While the inclusion of a question concerning conceptual overlap in the recent mapping reporting standard [7] should increase the attention this issue receives, to date only a very small number of studies e.g. [22–24] have explicitly recognised the problem.

3 Potential Consequences of Poor Conceptual Validity When Mapping

Different aspects of HRQoL will be included in mapped estimates from different condition-specific measures due to the differing conceptual overlap between the descriptive systems of pairs of instruments. This may result in systematic biases in the preferences for health that are lost and retained when mapped estimates are derived.

For example, the Multiple Sclerosis Impact Scale-29 (MSIS-29) [25, 26] is a patient-reported outcome measure that assesses the impact of multiple sclerosis (MS) on people's HRQoL. It has two subscales that assess the

impact of MS on physical functioning and psychological functioning, but does not include questions relating to mobility. The Multiple Sclerosis Walking Scale-12 (MSWS-12) [27] is also a patient-reported outcome measure that specifically assesses the impact of MS on people's mobility. Mapping from scores on the MSIS-29 to the EQ-5D will result in EQ-5D estimates that exclude, to a great extent, aspects of HRQoL pertaining to mobility, while mapping from scores on the MSWS-12 to the EQ-5D will result in EQ-5D estimates largely dominated by the impact of MS on mobility. The extent to which these mapped estimates might then be considered equivalent is highly questionable.

There is a focus in the literature on the effect of *adding* dimensional information when deriving health state utility values [28], however, there has been little, if any, consideration of the impact of *losing* dimensional information, as is the case with mapping. When mapping, content that is not included in the descriptive systems of *both* of the measures, e.g. a condition-specific measure and the EQ-5D, will be lost. In addition, when mapping, the salience of some of the dimensions may be altered, as illustrated in the example above with the MSIS-29 and MSWS-12. This could mean that the remaining health states may have been valued differently by the general population, i.e. the preference-based measure will have been valued based on *all* the included domains, but, in the process of mapping, much of this information may be lost, and the remaining health states may have been valued differently. As such, mapped estimates may not capture the genuine preferences of the general population.

4 Improving the Conceptual Validity of Mapping Algorithms

The assessment and treatment of conceptual validity and information loss in mapping is at an embryonic stage. The approaches to dealing with validity in mapping have focused almost entirely on the statistical predictive nature of mapping functions. This leaves a gap in our understanding of the mapping process as predictive accuracy can only tell us how closely scores are related, not whether they are assessing the same preferences for health. The following are preliminary suggestions for how conceptual validity might be assessed in the mapping process.

4.1 Response Mapping

Mapping is conducted in two main ways. The most common form involves scores on the starting measure being regressed directly onto health state utility values (e.g. EQ-5D index scores). The second approach uses a two-stage

process. Scores on the starting measure are firstly regressed onto responses on the health status measure (e.g. EQ-5D dimension scores), and health state utility values, usually from general population samples, are thereafter assigned to them.

This latter approach, termed response mapping [29], appears to offer no clear statistical advantage over mapping directly to health state utility values. It seems to work better at predicting individual scores and score ranges, but indications are that it generates larger errors across groups [30–35]. However, as Parkin et al. [36] describe, using EQ-5D index data, which are derived from people's reports of health status to which health state utility values from the general population are applied, means that any statistical analysis with the EQ-5D index is affected not just by variations in scores due to the sample but also by variations in scores due to the values given by the general population. When mapping, this serves to add unidentified complexity to the relationship between scores on the starting measure and EQ-5D index scores. As such, it may be argued that response mapping offers greater conceptual clarity regarding the relationship between measures, what is being mapped to what, and what is lost and retained in the mapping procedure. Also, unless response mapping is used, the process of mapping does not directly relate conceptual dimensions on the starting measure to conceptual dimensions on the target measure. Rather, scores on conceptual dimensions on the starting measure are related to general population preferences for degrees of the conceptual dimensions on the target measure.

4.2 Decision Validity

Without considering conceptual validity, it is not possible to assess whether resource allocation decisions based on such estimates are valid. The validity of an individual instrument has traditionally been measured according to a set of established criteria and, in particular, four key criteria: face, content, criterion and construct validity [37]. This differs from how validity needs to be assessed as part of the development of mapping functions, where it will be necessary to establish a method for assessing the validity of the process of conversion from one instrument to another. This may use the established definitions of validity but must also consider if there are other criteria by which this function can be judged.

The concept of decision validity [38] may offer such an additional criterion. Decision validity refers to the degree of certainty that the decision made is the correct one, given the available information. Decision validity then is the ultimate arbiter for any mapping function; there is a direct relationship between the robustness of mapped utility estimates and the validity of any decision informed by

those estimates. Thus, if the conceptual uncertainty surrounding a decision falls within an acceptable range given the currently available information, it can be concluded that the decision is valid and therefore that the mapping function is valid.

For establishing *conceptual* decision validity, there must be a *prima facie* case that a function that maps between any two given instruments satisfies the set of conditions outlined below. Analysts should assess the following:

- (a) the degree to which the two instruments measure the same or similar concepts (face validity);
- (b) the degree to which the two instruments measure the same hypothetical constructs (construct validity);
- (c) the extent that scores from the source instrument correlate with those from the target (or gold standard) instrument (criterion validity).

If the analyst is satisfied by the expected degree of conceptual coherence between the two instruments, there is a *prima facie* case that a mapping function may potentially lead to a valid decision. There is currently no specific guidance on what degree of conceptual overlap is required to proceed with the production of a mapping algorithm. Further research is needed to inform guidance on appropriate methods for assessing conceptual overlap and the cut-offs at which a lack of coherence is liable to result in erroneous resource allocation decisions (compared with the application of directly derived health utility values).

This is, of itself, not sufficient. A method is needed to be able to establish the nature of *what* is lost when a mapping algorithm is used, not just *how much* is lost. For example, if aspects of HRQoL pertaining to mobility are lost in the mapping process because there is no assessment relating to mobility in the starting measure, and therefore no overlap on this domain between the starting measure and the target measure, it seems important to know that the health state utility value estimates will contain very little information relating to mobility and preferences for mobility-affected health states. Assessing what content is retained and what content is lost in a mapping algorithm might be addressed by starting with a conceptual map of the construct being mapped from (e.g. on the basis of good-quality qualitative research). For example, if wishing to map from the Fatigue Severity Scale to the EQ-5D, the conceptual overlap assessment could begin by setting out a framework of the areas of impact of fatigue on people's lives. The descriptive content of the preference-based measure being mapped to, e.g. the EQ-5D, could then be assessed on an item-by-item basis in relation to this framework to identify areas of overlap and difference.

Analysts are encouraged to consider the conceptual validity of mapping in relation to their resource allocation decision, to consider what criteria might be used to assess

conceptual validity, to consider the possible consequences of using mapping algorithms that are not ‘conceptually valid’, and to sometimes not map because it is not conceptually valid to do so.

5 Conclusions

The conclusions drawn in this paper are at odds with NICE guidance suggesting the use of mapped estimates in health technology appraisals [4, 5]. The acceptance of mapping before key conceptual aspects of the methodology are established may lead to a series of unintended and unwelcome consequences; treatments may erroneously be approved or rejected or investment may be made on the basis of poor information about the quality of life of the intended population. As is shown in this article, current methods used for mapping are not known to be conceptually robust. Any allocation decisions taken based on current mapped estimates, even if by chance they are the ‘correct’ decisions, can currently only be described as valid in a post hoc manner; however, waiting until the decision has been implemented to determine whether it is a valid decision does not address the original problem. Until the methodology exists for determining the validity of decisions based on mapped estimates, mapping should not be used for estimating preferences, and the best option remains valuation through the use of an established preference-based instrument or valuation technique. We agree with the conclusions of McCabe et al. [8] that, in its current guise, there is a significant risk that mapping may be harmful to population health, and recommend that mapping research now focuses on developing criteria for when it is and is not conceptually valid to derive a mapping algorithm.

Author Contributions JR and AH each contributed to the development of the arguments made in the manuscript. JR developed the first draft of the manuscript. Both JR and AH contributed to further drafting of the manuscript and agreed the final version.

Compliance with Ethical Standards

Funding Annie Hawton acknowledges funding from the Pen-CLAHRC National Institute for Health Research Collaborations for Leadership in Applied Health Research and Care. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Conflict of interest Jeff Round and Annie Hawton have no conflicts of interest to report.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any

medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Dakin H. HERC database of mapping studies, Version 1.0 (last updated: 14th January 2013). Available at: <http://www.herc.ox.ac.uk/downloads/mappingdatabase>.
2. EuroQol Group. EQ-5D user guide. Rotterdam: The EuroQol Group; 1996.
3. Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–25.
4. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2008. Manchester: National Institute for Health and Care Excellence; 2008.
5. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Manchester: National Institute for Health and Care Excellence; 2013.
6. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Med Decis Mak*. 2008;28(1):66–89. doi:10.1177/0272989X07309642.
7. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. *Pharmacoeconomics*. 2015;33(10):985–91.
8. McCabe C, Edlin R, Meads D, Brown C, Kharroubi S. Constructing indirect utility models: some observations on the principles and practice of mapping to obtain health state utilities. *Pharmacoeconomics*. 2013;31(8):635–41.
9. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, et al. The maps reporting statement for studies mapping onto generic preference-based outcome measures. *Value Health*. 2015;18(7):A715–6.
10. Longworth L, Rowen D. NICE DSU Technical Support Document 10. The use of mapping methods to estimate health state utility values. Available at: <http://www.nicedsu.org.uk>.
11. Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. HEDS Discussion Paper 12. Sheffield: University of Sheffield; 2012.
12. Lu G, Brazier J, Ades A. Mapping from disease-specific to generic health-related quality-of-life scales: a common factor model. *Value Health*. 2013;16(1):177–84.
13. Brazier J, Roberts J, Deverill M. The estimation of a preference-based single index measure for health from the SF-36. *J Health Econ*. 2002;21:271–92.
14. Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–9.
15. Skevington SM. Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. *Pain*. 1998;76(3):395–406.
16. Juniper E, Buist S, Cox F, Ferrie P, King D. Validation of a standardised version of the Asthma Quality of Life Questionnaire. *Chest*. 1999;115:1265–70.
17. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.
18. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, et al. Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR Good

- Practices for Outcomes Research Task Force Report. *Value Health*. 2017;20(1):18–27.
19. Longworth L, Buxton M, Sculpher M, Smith D. Estimating utility data from clinical indicators for patients with stable angina. *Eur J Health Econ*. 2005;6:347–53.
 20. Campeau L. Letter: grading of angina pectoris. *Circulation*. 1976;54(3):522–3.
 21. Fredriksson T, Pettersson U. Severe psoriasis: oral therapy with a new retinoid. *Dermatologica*. 1978;157(4):238–44.
 22. Round J. Capturing information loss in estimates of uncertainty that arise from mapping algorithms. Aberdeen: Health Economists' Study Group; 2008.
 23. Hawton A, Green C, Telford C, Wright D, Zajicek J. The use of multiple sclerosis condition-specific measures to inform health policy decision-making: mapping from the MSWS-12 to the EQ-5D. *Mult Scler J*. 2012;18(6):853–61.
 24. Hawton A, Green C, Telford C, Zajicek J, Wright D. Using the Multiple Sclerosis Impact Scale to estimate health state utility values: mapping from the MSIS-29, Version 2, to the EQ-5D and the SF-6D. *Value Health*. 2012;15:1084–91.
 25. Hobart J, Lamping D, Fitzpatrick R, Riaz A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29). A new patient-based outcome measure. *Brain*. 2001;124:962–73.
 26. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix–x, 1–177.
 27. Hobart J, Riaz A, Lamping D, Fitzpatrick R, Thompson A. Measuring the impact of MS on walking ability: the 12-item MS Walking Scale (MSWS-12). *Neurology*. 2003;60:31–6.
 28. Longworth L, Yang Y, Rowen D, Tsuchiya A, Young T, Brazier J. Development and valuation of a vision bolt-on to EQ-5D. Exeter: University of Exeter; 2013.
 29. Gray A, Rivero-Arias O, Clarke P. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Mak*. 2006;26:18–29.
 30. Edlin R, Tsuchiya A, Brazier J. Mapping the Minnesota Living with Heart Failure Questionnaire to the EQ-5D index. Sheffield: University of Sheffield; 2002.
 31. Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition specific instruments: converting AQLQ into EQ-5D indices. HEDS discussion paper. Sheffield: University of Sheffield; 2002.
 32. Rivero-Arias O, Ouellet M, Gray A, Wolstenholme J, Rothwell P, Luengo-Fernandez R. Mapping the modified Rankin scale (mRS) measurement into the generic EuroQol (EQ-5D) health outcome. *Med Decis Mak*. 2010;30(3):341–54.
 33. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res*. 2013;22(3):683–94.
 34. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes*. 2009;7:27.
 35. Chuang L, Kind P. Converting the SF-12 into the EQ-5D: an empirical comparison of methodologies. *Pharmacoeconomics*. 2009;27(6):491–505.
 36. Parkin D, Rice N, Devlin N. Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? *Med Decis Mak*. 2010;30(5):556–65.
 37. Streiner D, Norman G. *Health measurement scales*. 3rd ed. Oxford: Oxford University Press; 2003.
 38. Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. *Health Econ*. 2004;11(1):1–8.